Vulnerabilities of transcriptome-wide association studies

Michael Wainberg[1], Nasa Sinnott-Armstrong[2], David A Knowles[2,3], David Golan[2], Raili Ermel[4], Arno Ruusalepp[4,5], Thomas Quertermous[6], Ke Hao[7], Johan LM Björkegren[5,7,8,9,], Manuel A Rivas[10,*], Anshul Kundaje[1,2,*]*

Transcriptome-wide association studies (TWAS) integrate GWAS and expression quantitative trait locus (eQTL) datasets to discover candidate causal gene-trait associations. We integrate multi-tissue expression panels and summary GWAS for LDL cholesterol and Crohn's disease to show that TWAS are highly vulnerable to discovering non-causal genes, because variants at a single GWAS hit locus are often eQTLs for multiple genes. TWAS exhibit acute instability when the tissue of the expression panel is changed: candidate causal genes that are TWAS hits in one tissue are usually no longer hits in another, due to lack of expression or strong eQTLs, while non-causal genes at the same loci remain. While TWAS is statistically valid when used as a weighted burden test to identify trait-associated loci, it is invalid to interpret TWAS associations as causal genes because the false discovery rate for TWAS causal gene discovery is not only high, but unquantifiable. More broadly, our results showcase limitations of using expression variation across individuals to determine causal genes at GWAS loci.

[1] *Department of Computer Science, Stanford University, Stanford, California, USA.*
[2] *Department of Genetics, Stanford University, Stanford, California, USA.*
[3] *Department of Radiology, Stanford University, Stanford, California, USA.*
[4] *Department of Cardiac Surgery, Tartu University Hospital, Tartu, Estonia.*
[5] *Clinical Gene Networks AB, Stockholm, Sweden.*
[6] *Division of Cardiovascular Medicine, Stanford University, Stanford, California, USA.*
[7] *Department of Genetics & Genomic Sciences, Institute of Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, New York, USA.*
[8] *Department of Pathophysiology, Institute of Biomedicine and Translational Medicine, University of Tartu, Tartu, Estonia.*
[9] *Division of Vascular Biology, Department of Medical Biochemistry and Biophysics, Karolinska Institutet, Stockholm, Sweden.*
[10] *Department of Biomedical Data Science, Stanford University, Stanford, California, USA.*
[*] *Correspondence to: Johan Björkegren (johan.bjorkegren@mssm.edu) or Manuel Rivas (mrivas@stanford.edu) or Anshul Kundaje (akundaje@stanford.edu).*

Introduction

Transcriptome-wide association studies (TWAS) are a recent family of methods that leverage expression reference panels (eQTL cohorts with expression and genotype data) to discover associations in GWAS datasets[1,2]. TWAS begin by building predictive models of gene expression from allele counts (typically using variants within a window of 500 kb or 1 MB around the gene), then use these models to predict expression for each individual in the GWAS cohort and associate this predicted expression with the trait (Fig. 1).

TWAS have garnered substantial interest within the human genetics community and TWAS have subsequently been conducted for a wide variety of traits and tissues[3]. A key reason for the appeal of TWAS is the promise that gene-disease associations represent likely causal genes, although both papers are careful not to claim causality with absolute certainty.

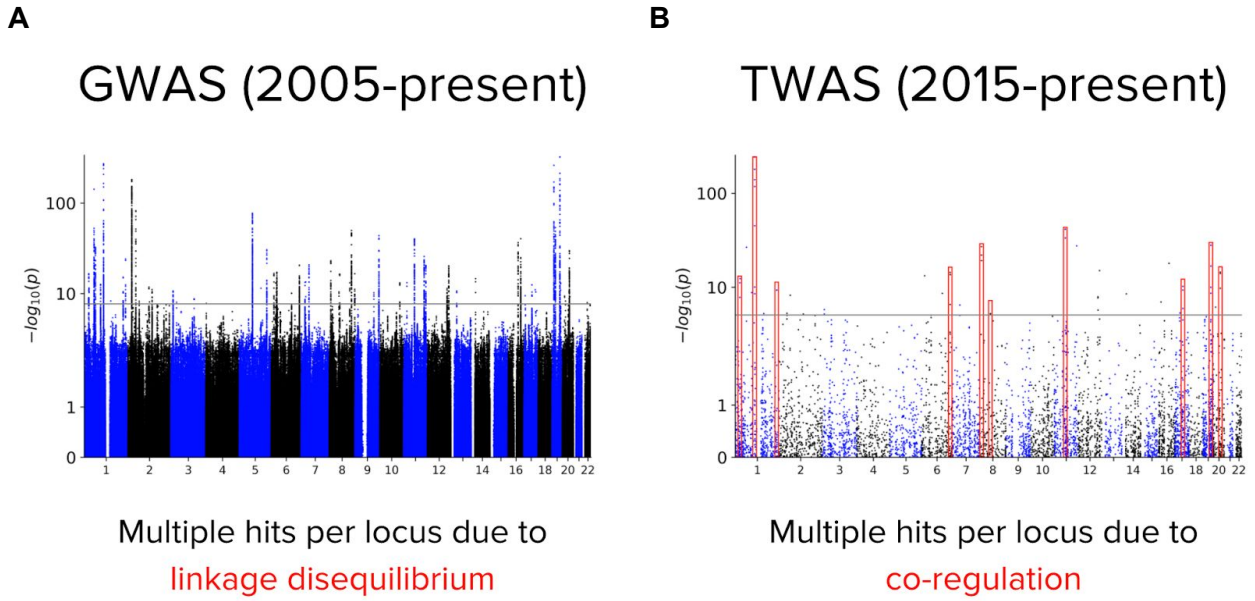
Alternatively, TWAS can be interpreted as a weighted burden test. All existing TWAS methods use a linear expression model, which means that TWAS is equivalent to testing a linear combination of variants against the phenotype, where the weights of the linear combination have been chosen based on how much the variant is predicted to contribute to expression variation across individuals in the reference panel. The goal of a weighted burden test is to increase power relative to single-variant testing (GWAS).

Results

TWAS loci frequently contain multiple associated genes

It is well known that GWAS rarely identifies single variant-trait associations, but instead identifies blocks of associated variants in linkage disequilibrium (LD) with each other (Fig. 1a). Unexpectedly, TWAS also frequently identifies multiple hit genes per locus (Fig. 1b), a phenomenon observed previously[4].

To explore this phenomenon, we performed TWAS in two traits and two tissues with Fusion, using GWAS summary statistics for LDL cholesterol[5] and Crohn's disease[6] and the 522 liver and 447 whole blood expression samples from the STARNET cohort[7] (Fig. S2, Online Methods). We grouped hit genes within 2.5 MB and found that while some loci contained only a single hit gene, many contained two, three, four or even up to eleven (Fig. S3).



C

Scenario	Est. % of non-causal hit genes	Case study locus
Correlated expression	20% ($r^2 \geq 0.2$)	*SORT1* (LDL, liver)
Correlated predicted expression	75% ($r^2 \geq 0.2$)	*IRF2BP2* (LDL, liver)
Expression models share variants	69% (≥ 1 shared)	*NOD2* (Crohn's, whole blood)

Figure 1: TWAS, like GWAS, frequently has multiple hits per locus. (a), (b) Manhattan plots of GWAS and TWAS for LDL cholesterol using GWAS summary statistics from the Global Lipids Genetics Consortium and liver expression from the STARNET cohort (see Methods). GWAS has multiple hits per locus due to linkage disequilibrium, and TWAS due to co-regulation, as we explore in the paper. Clusters of multiple adjacent TWAS hit genes are highlighted in red. (c) Three scenarios where co-regulation can lead to multiple hits per locus, and the estimated percent of non-causal hit genes subject to each scenario; each scenario is presented in a case study later in the paper (a fourth scenario is presented in Fig. 5d). To estimate the percentages, we group hits into 2.5 MB clumps and make the approximation that genes that are not the top hit in multi-hit clumps are non-causal; we then calculate the percent of these genes with total or predicted expression $r^2 \geq 0.2$ or ≥ 1 shared variant with the top hit in their block, aggregating genes across the LDL/liver and Crohn's/whole blood TWAS. The full distributions of total and predicted expression correlations and number of shared variants are shown in Fig. S1, separated by study.

Correlated expression across individuals may lead to non-causal TWAS hit genes

We wondered whether co-regulation could be responsible for multi-hit loci. The conventional way co-regulation is measured is by correlating the expression of a pair of genes across individuals in an expression cohort. Do genes that have correlated expression with a strong TWAS hit also tend to be TWAS hits (Fig. 5a)? We analyzed the *SORT1* locus in LDL/Liver (TWAS $p < 1 \times 10^{-243}$; Fig. 2a) which represents the strongest hit locus across all four TWAS. *SORT1* has strong evidence of causality, though not without some controversy over the precise mechanism: in mouse models, overexpression of *SORT1* in liver reduced plasma LDL levels and siRNA knockdown increased plasma LDL levels[8,9], though in other studies deletion of *SORT1* counter-intuitively reduced, rather than increased, atherosclerosis in mice without affecting plasma LDL levels[10,11,12].

The *SORT1* locus contains 8 other TWAS hit genes besides *SORT1*, and their TWAS *p*-values are highly related to their expression correlation with *SORT1* (Spearman = 0.75; Fig. 2b). Given that *SORT1* has strong evidence of causality, and that other genes at the locus lack strong literature evidence, the most parsimonious explanation is that most or all of the other genes are non-causal and are only hits due to their correlation with *SORT1*.

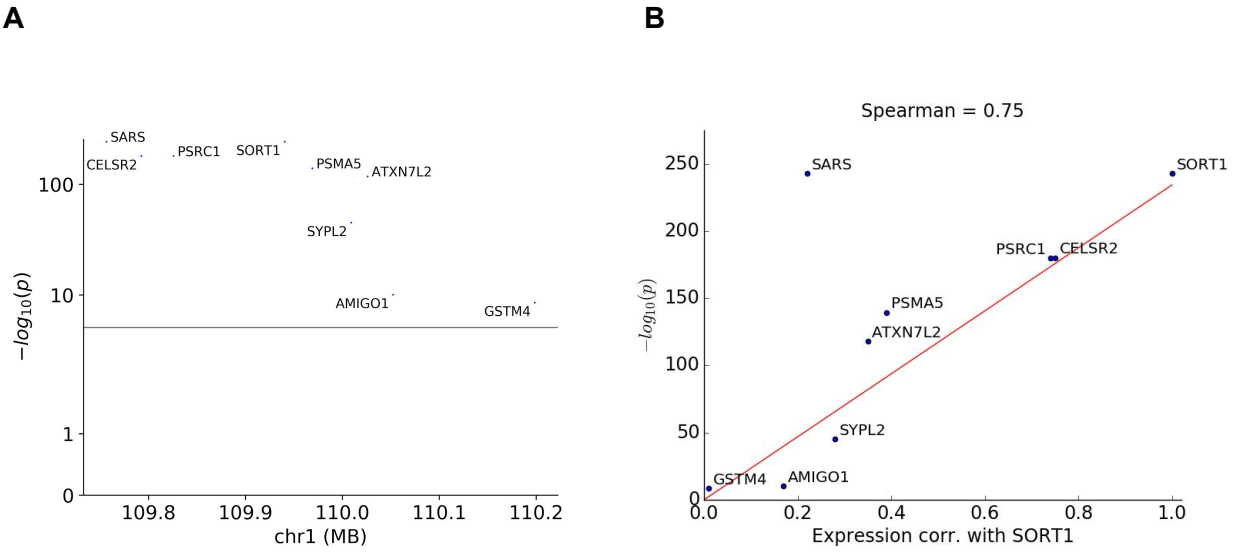


Figure 2: Co-regulation strongly predicts TWAS hit strength at the *SORT1* locus. a) TWAS Manhattan plot of the *SORT1* locus. b) Expression correlation with *SORT1* versus TWAS *p*-value, for each gene in the *SORT1* locus.

Correlated predicted expression is sufficient for non-causal hits even without correlated total expression

However, expression correlation is not the whole story: after all, TWAS tests for association with predicted expression, not total expression. Total expression includes genetic,environmental and technical components, and the genetic component of expression includes contributions

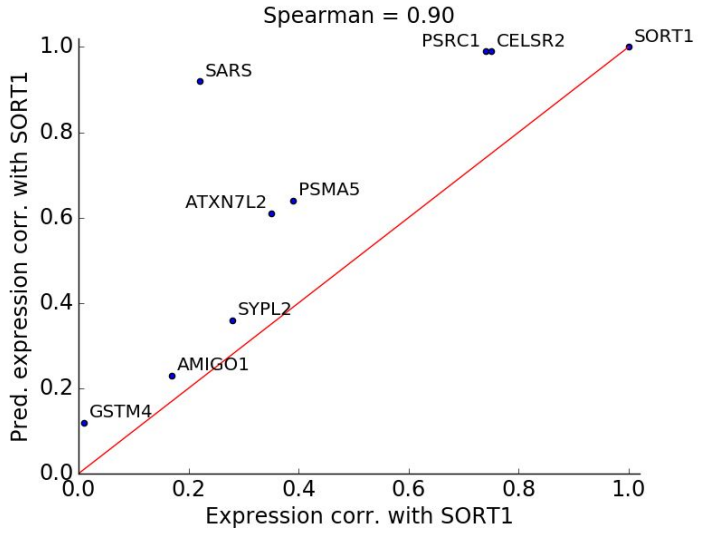
from common *cis* eQTLs (the only component reliably detectable in current TWAS methods), rare *cis* eQTLs, and *trans* eQTLs. Predicted expression likely only represents a small component of the GWAS individuals' total expression: a large-scale twin study[13] found that common *cis* eQTLs explain only about 10% of genetic variance in gene expression.

While predicted expression correlations between genes at the same locus are often similar to total expression correlations, they are generally slightly higher, and sometimes substantially so (Fig. 3a, Fig. S4). It is reasonable to expect nearby genes to be more tightly co-regulated at the level of *cis* expression than at the level of total expression, since even if distinct *trans* and environmental effects act on the two genes, they do at least share the same *cis* sequence context.

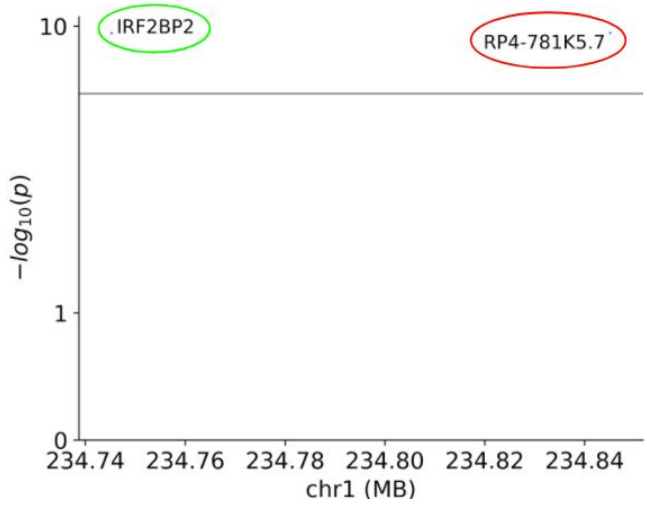
Predicted expression correlation may lead to non-causal hits even for genes with low total expression correlation (Fig. 5b). For instance, *SARS* is the main outlier in Fig. 2b because, despite having a similar TWAS *p*-value to *SORT1*, it has an unexpectedly low total expression correlation of approximately 0.2; yet it is still a strong hit because of its high predicted expression correlation of approximately 0.9 (Fig. 3a).

Another example is the *IRF2BP2* locus in LDL/liver (Fig. 3b), where *RP4-781K5.7* is a likely non-causal hit due to predicted expression correlation with *IRF2BP2*, a gene encoding an inflammation-suppressing regulatory factor with strong evidence of causality from mouse models, at least at the level of atherosclerosis[14]. While there is almost no correlation in total expression between the two genes (Pearson = -0.02), *IRF2BP2*'s expression model includes a GWAS hit variant, rs556107, with a negative weight while *RP4-781K5.7*'s includes the same variant, as well as two other linked variants, with positive weights (Fig. 3c), resulting in almost perfectly anti-correlated predicted expression between the two genes (Pearson = -0.94).

A



B



C

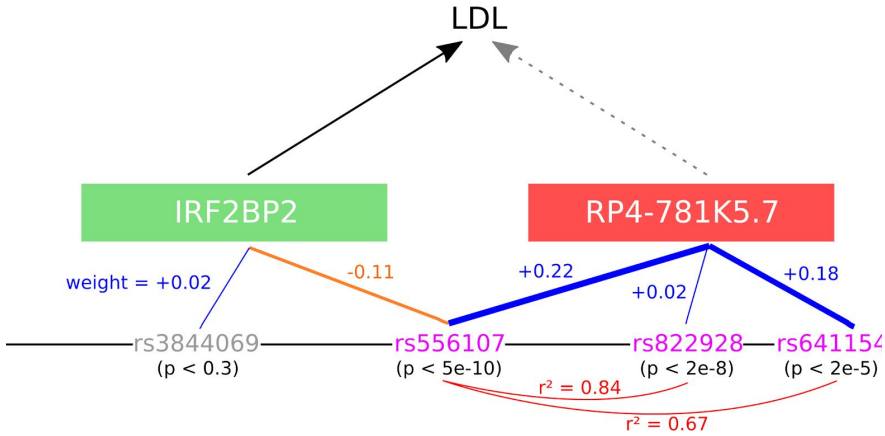


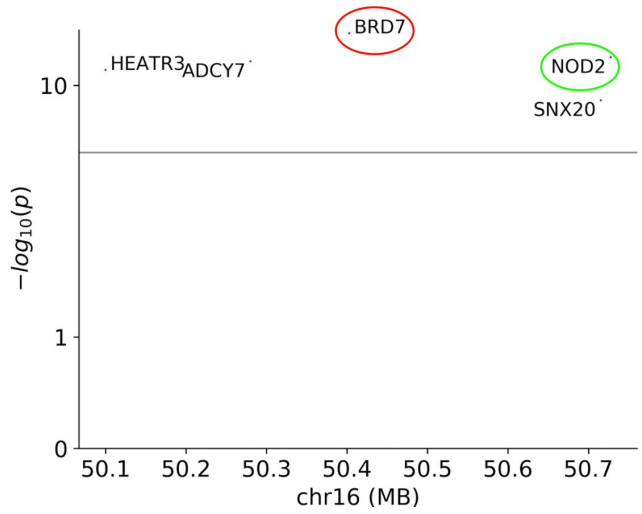
Figure 3: Correlated predicted expression can cause non-causal hits even in the absence of correlated total expression. a) For nearby genes, predicted expression correlations tend to be higher than total expression correlations, e.g. at the *SORT1* locus. b) TWAS Manhattan plot of the *IRF2BP2* locus, where *RP4-781K5.7* is a likely non-causal hit due to predicted expression correlation with *IRF2BP2*. c) Details of the two genes' expression models: a line between a variant's rs number and a gene indicates the variant is included in the gene's expression model with either a positive weight (blue) or negative weight (orange), with the thickness of the line increasing with the magnitude of the weight; red arcs indicate LD. Pink rs numbers are GWAS hits (genome-wide-significant or sub-significant) while gray rs numbers are not.

<u>Shared GWAS variants can cause non-causal hits even without correlated predicted expression</u>

More generally, pairs of genes may share GWAS variants in their models even if they have low predicted expression correlation, since other variants that are distinct between the models may "dilute" the correlation (Fig. 5c). For instance, at the *NOD2* locus for Crohn's/whole blood,

NOD2 is a known causal gene[15,16], but 4 other genes are also TWAS hits (Fig. 4a), none with strong evidence of causality (though rare variants in one gene, *ADCY7*, have been associated with the closely related disease ulcerative colitis but not Crohn's[17]). The TWAS model for the strongest hit at the locus, *BRD7*, puts most of its weight on rs1872691, which is also the strongest GWAS variant in *NOD2*'s model (Fig. 4b). However, the *NOD2* model puts most of its weight on two other variants, rs7202124 and rs1981760, which are slightly weaker GWAS hits. The result is that even though *BRD7* appears to be a non-causal hit because of co-regulation with *NOD2*, the overall predicted expression correlation between the two genes is very low (-0.03), as is the total expression correlation (0.05).

A



B

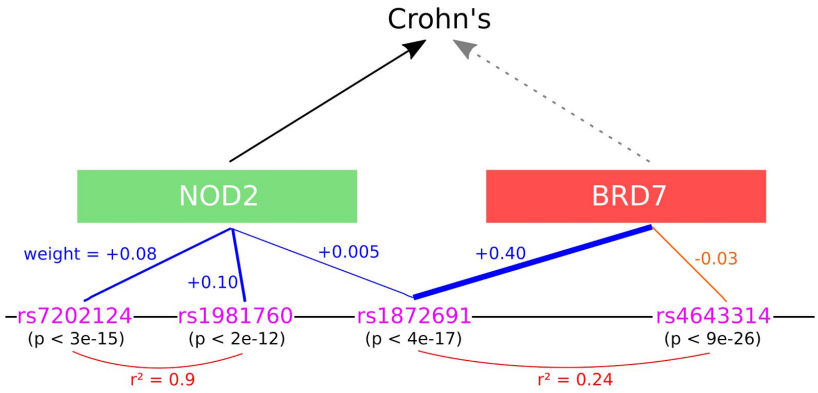
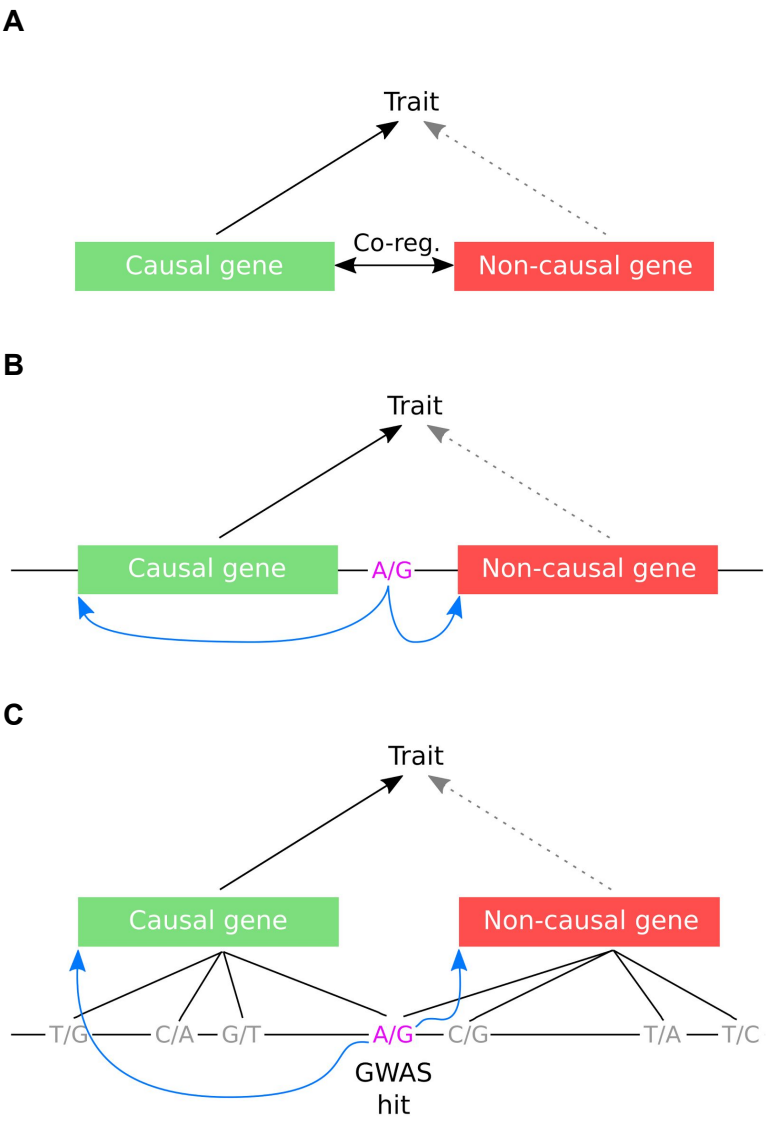


Figure 4: Sharing of GWAS variants between expression models can contribute to non-causal hits even without correlated predicted expression. a) TWAS Manhattan plot of the *NOD2* locus. b) Details of the expression models of *NOD2* and *BRD7*: as in Fig. 3, a line between a variant's rs number and a gene indicates the variant is included in the gene's expression model with either a positive weight (blue) or negative weight (orange), with the thickness of the line increasing with the magnitude of the weight; red arcs indicate LD.

In the most general case, models need not even share the same GWAS variants for there to be non-causal hits (Fig. 5d). For instance, the other two variants in *NOD2*'s model are neither shared nor in strong LD with of the variants in *BRD7*'s model (Fig. 4b). Under the assumption that *NOD2* is the only causal gene at the locus, this suggests that these variants exert their GWAS effects via *NOD2* and also happen to co-regulate *BRD7*, but the *NOD2* expression model incorrectly fails to include them (a false negative). This type of scenario might occur even without any false negatives in the expression modeling, e.g. if the two *NOD2* variants (or variants in LD) deleteriously affected the coding sequence of *NOD2* as well as regulating *BRD7*. Co-regulation is difficult to quantify, let alone correct for.



D

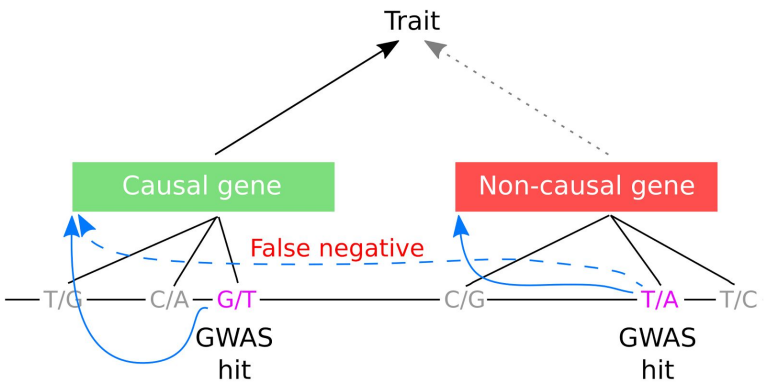


Figure 5: Co-regulation scenarios in TWAS that may lead to non-causal hits, from least to most general. a) Correlated expression across individuals: the causal gene has correlated total expression with another gene, which may become a non-causal TWAS hit. b) Correlated predicted expression across individuals: even if total expression correlation is low, predicted expression correlation may be high if the same variants (or variants in LD) regulate both genes and are included in both models. c) Sharing of GWAS hits: even if the two genes' models include largely distinct variants and predicted expression correlation is low, only a single shared GWAS hit variant (or variant in LD) is necessary for both genes to be TWAS hits. d) Both models include distinct GWAS hits: in the most general case, the GWAS hits driving the signal at the two genes may not be in LD with each other, for instance if the non-causal gene's GWAS hit happens to regulate the causal gene as well but this connection is missed by the expression modeling (a false negative), or if the causal gene's GWAS hit acts via a coding mechanism (not shown).

<u>Using expression from less related tissues substantially worsens the effects of co-regulation</u>

So far, our TWAS case studies have used expression from tissues with a clear mechanistic relationship to the trait: liver for LDL and whole blood for Crohn's. What if we swap these tissues (liver for Crohn's and whole blood for LDL), so that we are using tissues without a clear mechanistic relationship? It is well-known that the architecture of eQTLs differs substantially across tissues: even among strong eQTLs in GTEx ($p \sim 1 \times 10^{-10}$), one quarter switch which gene they are most significantly associated with across tissues[18].

We manually curated causal genes from the literature at 9 LDL/liver and 4 Crohn's/whole blood multi-hit TWAS loci and looked at how their hit strengths changed when swapping tissues (Fig. 6). Strikingly, almost every candidate causal gene (9 of 11 for LDL and 5 of 6 for Crohn's) was no longer a hit in the "opposite" tissue, either because they were not sufficiently expressed (N = 4: *PPARG*, *LPA*, *LPIN3*, *SLC22A4*) or because they did not have sufficiently heritable *cis* expression, according to a likelihood ratio test, to be tested by Fusion (N = 10: *SORT1*, *IRF2BP2*, *TNKS*, *FADS3*, *ALDH2*, *KPNB1*, *SLC22A5*, *IRF1*, *CARD9*, *STAT3*).

Worse, 15 other genes at the same loci were still hits (8 in LDL/whole blood and 7 in Crohn's/liver), and 5 were even strong hits with $p < 1 \times 10^{-20}$. This suggests that the strategy of conducting TWAS in a tissue that is sub-optimal for the trait being examined (e.g. whole blood, lymphoblastoid cell lines), just because that tissue happens to have a large expression

reference panel, is especially problematic because many hit loci may contain only non-causal genes and the causal gene may not even be included in the list of hits.

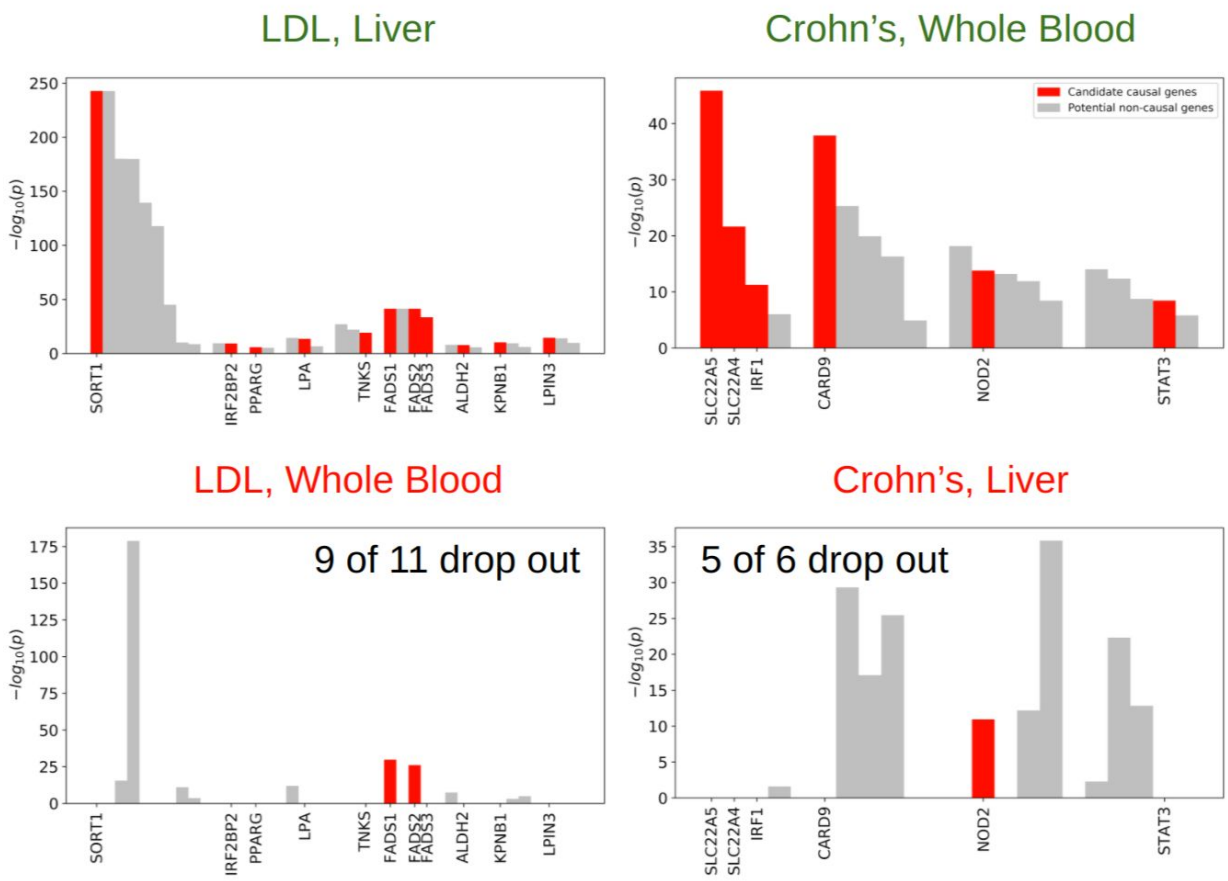


Figure 6: Most candidate causal genes drop out when switching to a tissue with a less clear mechanistic relationship to the trait, due to lack of sufficient expression or sufficiently heritable expression. TWAS *p*-values at 9 LDL/liver and 4 Crohn's/whole blood multi-hit loci, when using expression from tissues with a clear (top row) and less clear or absent (bottom row) mechanistic relationship to the trait. Candidate causal genes are labeled and colored in red.

Discussion

We have shown that it is invalid to interpret TWAS associations as causal genes, since it is highly vulnerable to non-causal gene-trait associations, intuitively because GWAS hits may be eQTLs for multiple genes. However, the ways in which co-regulation may lead to non-causal hits in TWAS are multi-faceted; co-regulation is hard to quantify, let alone correct for. The problem is particularly acute when using expression from tissues without a clear mechanistic relationship to the trait. It is valid to use TWAS as a weighted burden test, where the goal is not to identify causal genes but merely discover associated loci. While we appreciate the value of TWAS as a gene-based prioritization method, our analysis indicates that using TWAS for causal gene identification has strong interpretational limitations.

Is it possible, despite the limitations of TWAS, to somehow perform statistical fine-mapping and determine the causal gene or genes? We believe that it is not, even in principle. This is because predicted expression only imperfectly captures *cis* expression, the component of expression driven by variants near the gene; there are sources of both variance and bias in the expression modeling. The main source of variance is the finite size of the reference panel, although this can be mitigated with resampling methods. More problematically, the choice of tissue, cell type composition and quantification of the expression panel can all introduce bias. We have shown that using a tissue with a less clear mechanistic relationship to the trait hinders the ability to detect most candidate causal genes. Yet diseases rarely act through a single tissue: different genes may be causal in different tissues, so even using a tissue where most genes are causal may introduce bias for the remaining genes that are causal in a different tissue. Furthermore, most expression panels are gathered for tissues, not cell types, and genes may only be causal for a single cell type within a tissue: for instance, a study that identified *IRX3* and *IRX5* as causal genes at the *FTO* locus found genotype-expression associations in primary preadipocytes, a minority of adipose cells, but not in whole adipose tissue[19]. There may be substantial cell type heterogeneity within and between samples (e.g. due to the presence of blood and immune cells, or genetically-driven differences in the relative proportions of cell types within a tissue), which can also introduce bias. It is impossible to quantify every source of bias. On the other hand, fine-mapping may provide improved prioritization of causal genes in practice, although this should be evaluated based on known causal genes rather than merely the degree of reduction of the number of causal candidate genes at the locus.

In our case studies, we have generally assumed that the single gene with substantial evidence of causality is the sole causal gene at the locus, with some exceptions where there are multiple candidates and the causal gene or genes are under debate (*FADS1-3*, *SLC22A4*/*5*/*IRF1*). While this is the most parsimonious explanation, it is possible that some loci harbor multiple causal genes. Indeed, under an omnigenic model of complex traits[20], every gene may be causal to some degree, though it is still problematic if TWAS identifies marginally causal genes as strong hits due to co-regulation. Furthermore, the expression of other genes at the locus may causally contribute to the expression of the causal gene, merely by being actively transcribed, even if the gene is non-coding or its protein product has no causal role[21].

The vulnerabilities we have identified in TWAS, co-regulation and tissue bias, also apply to other methods that integrate GWAS and expression data. Gene-trait association testing based on Mendelian Randomization (MR)[22,23,24] is vulnerable to non-causal hits because co-regulation, as a form of pleiotropy, violates one of the core assumptions of MR[25]. While the HEIDI test[22] is designed to correct MR in the case where the two genes have distinct, but linked, causal variants, it does not control for the case where the two genes share the same causal variant. GWAS-eQTL colocalization methods such as Sherlock[26], coloc[27,28], QTLMatch[29], eCaviar[30], enloc[31] and RTC[32] are also vulnerable to this phenomenon. The more tightly a pair of genes is co-regulated in *cis*, the more difficult it becomes to distinguish causality based on GWAS and expression data alone. Our results underscore the need for computational and experimental

methods that move beyond using expression variation across individuals to determine the causal genes at GWAS loci.

Methods

TWAS were performed with the Fusion software (https://github.com/gusevlab/fusion_twas/tree/9142723485b38610695cea4e7ebb508945ec006c), using default settings and also including polygenic risk score as a possible model during cross-validation in addition to BLUP, Lasso, and ElasticNet. Variants in the STARNET reference panel were filtered for quality control using PLINK[33] with the options "--maf 1e-10 --hwe 1e-6 midp --geno". STARNET expression was processed as described in the STARNET paper[7], including probabilistic estimation of expression residuals[34] (PEER) covariate correction. Because Fusion, to our knowledge, only supports training on PLINK version 1 hard-call genotype files and not genotype dosages, we trained expression models on only the variants both genotyped in STARNET and either genotyped or imputed in the GWAS, filtering out variants without matching strands between the GWAS and STARNET. Expression models were trained on all remaining variants within 500 kb of a gene's TSS, using Ensembl v87 TSS annotations for hg19[35]. Linkage disequilibrium and total and predicted expression correlations were calculated across individuals in STARNET. Code to replicate the post-TWAS analysis is available at https://github.com/Wainberg/Vulnerabilities_of_TWAS.

Acknowledgements

We gratefully acknowledge Jonathan Pritchard and Hua Tang for helpful discussions. This work was funded in part by the Natural Sciences and Engineering Research Council of Canada (NSERC) (grant PGSD3-476082-2015 to M.W.), Stanford Bio-X Bowes fellowship (to M.W.), Stanford Graduate Fellowship (to N.S.-A.), National Defense Science & Engineering Grant (to N.S.-A.), and NIH grant 1DP2OD022870 (to A.K.).

Author Contributions

M.W., M.A.R. and A.K. conceived of the study. M.W. performed analyses. N.S.-A., D.A.K. and D.G. provided intellectual input. R.E., A.R., T.Q., K.H. and J.L.M.B. provided assistance with the STARNET dataset. M.A.R. and A.K. supervised the study. M.W., M.A.R. and A.K. wrote the manuscript. All authors reviewed the manuscript.

Competing Financial Interests

The authors declare no competing financial interests.

References

1. Gamazon, E. R. *et al.* A gene-based association method for mapping traits using reference transcriptome data.

Nat. Genet. **47,** 1091–1098 (2015).

2. Gusev, A. *et al.* Integrative approaches for large-scale transcriptome-wide association studies. *Nat. Genet.* **48,** 245–252 (2016).
3. Mancuso, N. *et al.* Integrating Gene Expression with Summary Association Statistics to Identify Genes Associated with 30 Complex Traits. *Am. J. Hum. Genet.* **100,** 473–487 (2017).
4. Gusev, A. *et al.* Transcriptome-wide association study of schizophrenia and chromatin activity yields mechanistic disease insights. (2016). doi:10.1101/067355
5. Willer, C. J. *et al.* Discovery and refinement of loci associated with lipid levels. *Nat. Genet.* **45,** 1274–1283 (2013).
6. Liu, J. Z. *et al.* Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nat. Genet.* **47,** 979–986 (2015).
7. Franzén, O. *et al.* Cardiometabolic risk loci share downstream cis- and trans-gene regulation across tissues and diseases. *Science* **353,** 827–830 (2016).
8. Musunuru, K. *et al.* From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus. *Nature* **466,** 714–719 (2010).
9. Strong, A. *et al.* Hepatic sortilin regulates both apolipoprotein B secretion and LDL catabolism. *J. Clin. Invest.* **122,** 2807–2816 (2012).
10. Mortensen, M. B. *et al.* Targeting sortilin in immune cells reduces proinflammatory cytokines and atherosclerosis. *J. Clin. Invest.* **124,** 5317–5322 (2014).
11. Patel, K. M. *et al.* Macrophage sortilin promotes LDL uptake, foam cell formation, and atherosclerosis. *Circ. Res.* **116,** 789–796 (2015).
12. Westerterp, M. & Tall, A. R. SORTILIN: many headed hydra. *Circ. Res.* **116,** 764–766 (2015).
13. Grundberg, E. *et al.* Mapping cis- and trans-regulatory effects across multiple tissues in twins. *Nat. Genet.* **44,** 1084–1089 (2012).
14. Chen, H.-H. *et al.* IRF2BP2 Reduces Macrophage Inflammation and Susceptibility to Atherosclerosis. *Circ. Res.* **117,** 671–683 (2015).
15. Hugot, J. P. *et al.* Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature* **411,** 599–603 (2001).
16. Ogura, Y. *et al.* A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. *Nature* **411,**

603–606 (2001).

17. Luo, Y. *et al.* Exploring the genetic architecture of inflammatory bowel disease by whole-genome sequencing identifies association at ADCY7. *Nat. Genet.* **49,** 186–192 (2017).
18. GTEx Consortium *et al.* Genetic effects on gene expression across human tissues. *Nature* **550,** 204–213 (2017).
19. Claussnitzer, M. *et al.* FTO Obesity Variant Circuitry and Adipocyte Browning in Humans. *N. Engl. J. Med.* **373,** 895–907 (2015).
20. Boyle, E. A., Li, Y. I. & Pritchard, J. K. An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell* **169,** 1177–1186 (2017).
21. Engreitz, J. M. *et al.* Local regulation of gene expression by lncRNA promoters, transcription and splicing. *Nature* **539,** 452–455 (2016).
22. Zhu, Z. *et al.* Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat. Genet.* **48,** 481–487 (2016).
23. Hauberg, M. E. *et al.* Large-Scale Identification of Common Trait and Disease Variants Affecting Gene Expression. *Am. J. Hum. Genet.* **100,** 885–894 (2017).
24. Pavlides, J. M. W. *et al.* Predicting gene targets from integrative analyses of summary data from GWAS and eQTL studies for 28 human complex traits. *Genome Med.* **8,** 84 (2016).
25. Solovieff, N., Cotsapas, C., Lee, P. H., Purcell, S. M. & Smoller, J. W. Pleiotropy in complex traits: challenges and strategies. *Nat. Rev. Genet.* **14,** 483–495 (2013).
26. He, X. *et al.* Sherlock: detecting gene-disease associations by matching patterns of expression QTL and GWAS. *Am. J. Hum. Genet.* **92,** 667–680 (2013).
27. Wallace, C. *et al.* Statistical colocalization of monocyte gene expression and genetic risk variants for type 1 diabetes. *Hum. Mol. Genet.* **21,** 2815–2824 (2012).
28. Giambartolomei, C. *et al.* Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet.* **10,** e1004383 (2014).
29. Plagnol, V., Smyth, D. J., Todd, J. A. & Clayton, D. G. Statistical independence of the colocalized association signals for type 1 diabetes and RPS26 gene expression on chromosome 12q13. *Biostatistics* **10,** 327–334 (2009).
30. Hormozdiari, F. *et al.* Colocalization of GWAS and eQTL Signals Detects Target Genes. *Am. J. Hum. Genet.* **99,** 1245–1260 (2016).

31. Wen, X., Pique-Regi, R. & Luca, F. Integrating molecular QTL data into genome-wide genetic association analysis: Probabilistic assessment of enrichment and colocalization. *PLoS Genet.* **13,** e1006646 (2017).

32. Nica, A. C. *et al.* Candidate causal regulatory effects by integration of expression QTLs with complex trait genetic associations. *PLoS Genet.* **6,** e1000895 (2010).

33. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81,** 559–575 (2007).

34. Stegle, O., Parts, L., Piipari, M., Winn, J. & Durbin, R. Using probabilistic estimation of expression residuals (PEER) to obtain increased power and interpretability of gene expression analyses. *Nat. Protoc.* **7,** 500–507 (2012).

35. Aken, B. L. *et al.* The Ensembl gene annotation system. *Database* (2016). doi:10.1093/database/baw093

Supplementary Information

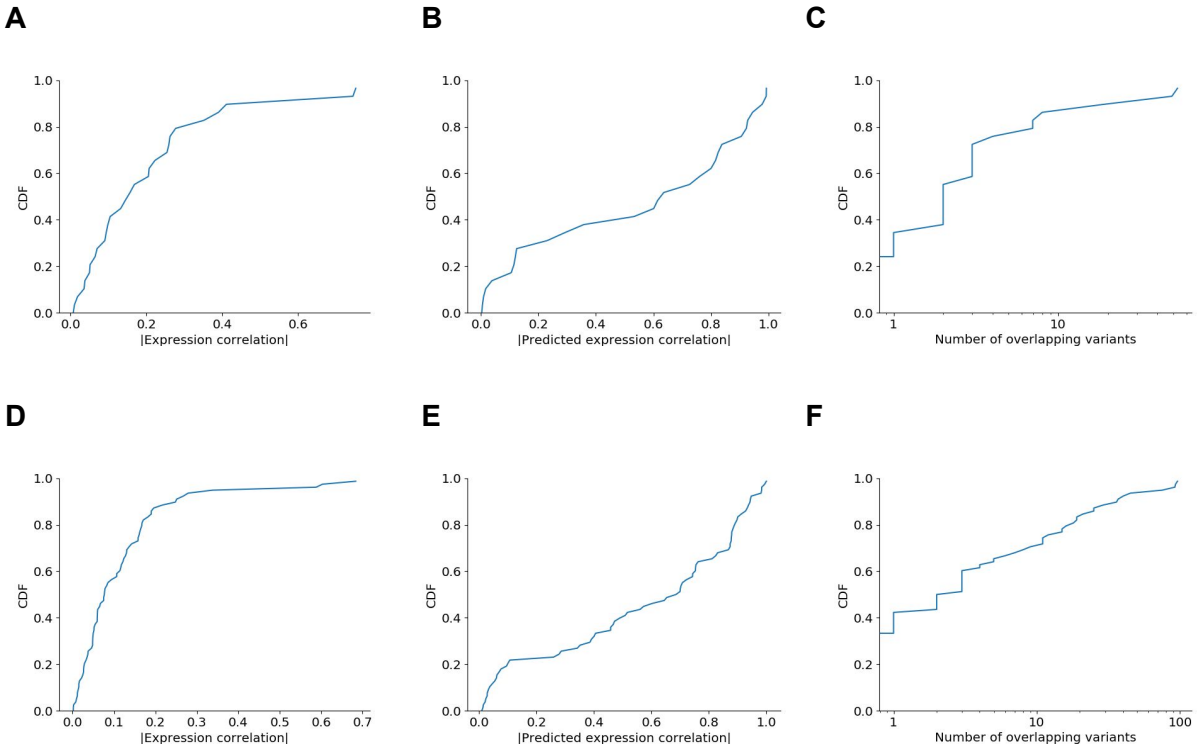


Figure S1: Distributions of co-regulation across putative non-causal genes in multi-hit TWAS loci. Since many multi-hit loci do not have a clear causal gene or have multiple plausible candidates, we make the approximation that only the most significant gene at each locus is causal. We then plot the cumulative distribution functions (CDFs) of (a, d) expression correlations, (b, e) predicted expression correlations and (c, f) number of shared variants between these most significant genes and all the other genes at their loci, separately for LDL/liver (a-c) and Crohn's/whole blood (d-f). To collapse these CDFs into a single estimate of the percent of affected non-causal genes (Fig. 1c), we combine genes across the two studies and threshold to correlation $r^2 \geq 0.2$, a threshold commonly used

for weak LD in GWAS, or ≥ 1 shared variant. Note that counting only exact sharing of variants does not account for LD, for simplicity.

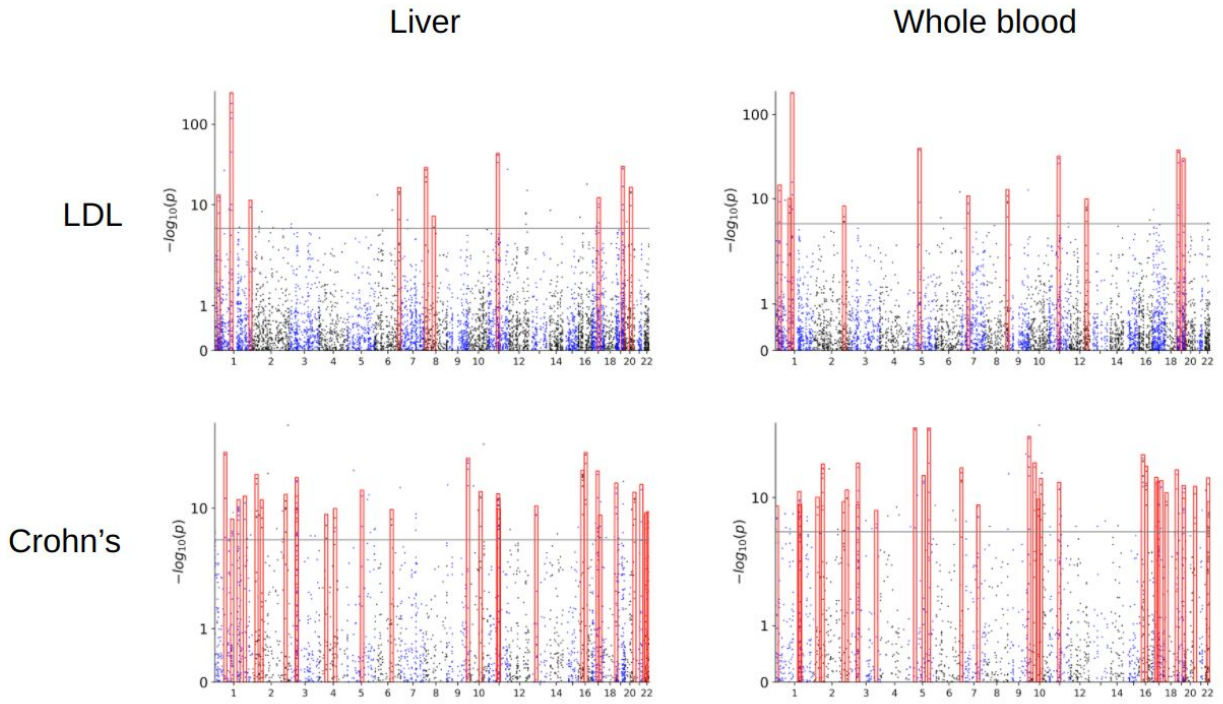


Figure S2: Manhattan plots of the 4 TWAS conducted in this study. As in Fig. 1, clusters of multiple adjacent TWAS hit genes are highlighted in red.

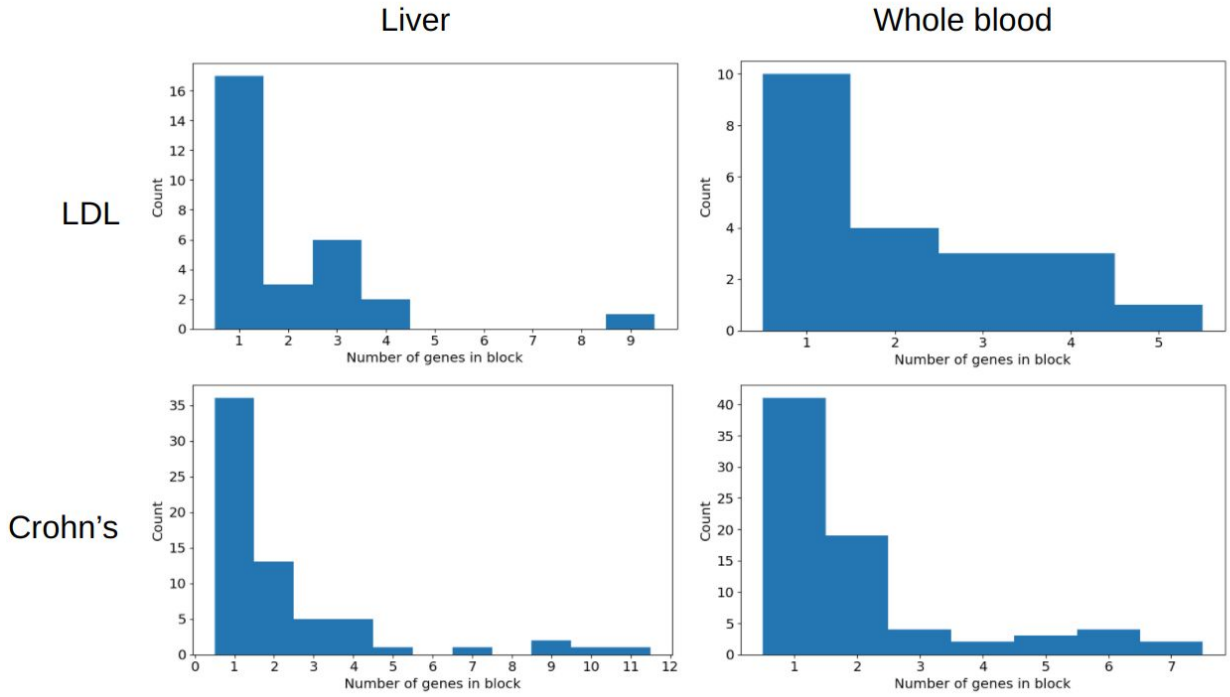


Figure S3: Number of TWAS hit genes per locus after 2.5-MB clumping.

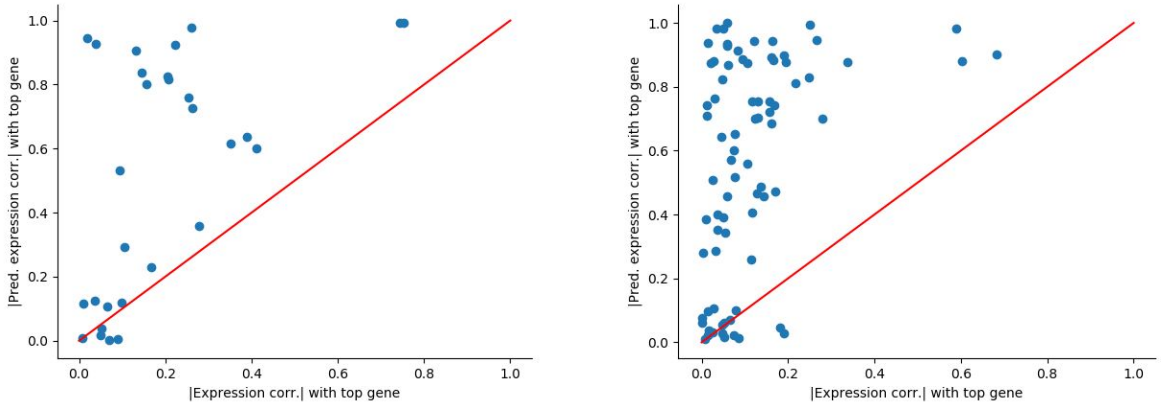


Figure S4: Total versus predicted expression correlation versus the top hit, for all genes in multi-hit blocks that are not the top hits. a) Liver, LDL. b) Crohn's, whole blood. Note that predicted expression correlation is generally higher than total expression correlation, as discussed in the Results section.